A Deep Learning-Based Multi-Sensor Data Fusion Method for Degradation Monitoring of Ball Screws

Li Zhang
Department of Mechanical Engineering,
Southwest Jiaotong University,
Chengdu 610031, China
Li_Zhang3836@163.com

Hongli Gao
Department of Mechanical Engineering,
Southwest Jiaotong University,
Chengdu 610031, China
hongli_gao@home.swjtu.edu.cn

***Abstract*--As ball screw has complex structure and long range of distribution, single signal collected by one sensor is difficult to express its condition fully and accurately. Multi-sensor data fusion usually has a better effect compared with single signal. Multi-sensor data fusion based on neural network(BP) is a commonly used multi-sensor data fusion method, but its application is limited by local optimum problem. Aiming at this problem, a multi-sensor data fusion method based on deep learning for ball screw is proposed in this paper. Deep learning, which consists of unsupervised learning and supervised learning, is the development and evolution of traditional neural network. It can effectively alleviate the optimization difficulty. Parallel superposition on frequency spectra of signals is directly done in the proposed deep learning-based multi-sensor data fusion method, and deep belief networks(DBN) are established by using fused data to adaptively mine available fault characteristics and automatically identify the degradation condition of ball screw. Test is designed to collect vibration signals of ball screw in 7 different degradation conditions by using 5 acceleration sensors installed on different places. The proposed fusion method is applied in identifying the degradation degree of ball screw in the test to demonstrate its efficacy. Finally, the multi-sensor data fusion based on neural network is also applied in degradation degree monitoring. The monitoring accuracy of deep learning-based multi-sensor data fusion is higher compared with that of neural network-based multi-sensor data fusion, which means the proposed method has more superiority.**

Keywords- multi-sensor data fusion; deep learning; ball screw; degradation monitoring

I. INTRODUCTION

Ball screw is an important component of drive system in machine tool to convert the rotary motion to translational motion. The performance degradation of ball screw will directly affect the machining accuracy and processing stability. Hence the degradation condition monitoring of ball screw is important. Effective condition monitoring could provide many benefits, such as increasing the system security and reliability, avoiding sudden failure which will result in big economic loss. As ball screw has complex structure and long range of distribution, multiple signals collected by multiple sensors can better express its degradation condition. Therefore, fusion technology is needed to fuse these signals collected by multiple sensors. The fused signals are used in degradation condition monitoring to estimate the degradation degree of ball screw.

The concept of information fusion has been put forward in 1973. Information fusion has experienced a long development process and is widely used in many fields like military, robot, image processing, speech recognition and some others. Neural network-based multi-sensor data fusion is commonly used in many studies. Kuo et al. [1] developed an estimation system through artificial neural networks and fuzzy logic to integrate multiple sensors to estimate on-line tool wear. There are also many researches applying information fusion in fault diagnosis and condition monitoring. Safizadeh and Latifi [2] fused vibration signals using fusion technology and load fused signals into bearing fault diagnosis to prove its effectiveness. Fused signals can express the equipment state more comprehensive and accurate.

However, local optimum problem limits the use of neural network-based multi-sensor data fusion. Unsupervised learning process is added on deep learning compared with traditional neural network. After the concept of deep learning is firstly proposed by Hinton and Salakhutdinov [3] in 2006, a lot of researches about deep learning have been done in many fields like computer vision, speech recognition and information retrieval over the past 10 years. There are also some studies have applied deep learning in fault diagnosis and condition monitoring [4-5]. Feng Jia [6] proposed a novel intelligent diagnosis method that DNNs are utilized to implement both fault feature extraction and intelligent diagnosis and has achieved good results. Deep learning can extract features from the frequency domain adaptively, which has become a new development direction of fault diagnosis and condition monitoring.

This paper proposed a degradation monitoring method for ball screw based on DBN and multi-sensor data fusion. The rest of the paper is organized as follow: Section II simply explains the theory of DBN and multi-sensor data fusion technology. Section III expresses the implementation steps of the proposed deep learning-based multi-sensor degradation monitoring method for ball screw in detail. In section IV, the effectiveness of the proposed method is validated through the

978-1-5090-2778-1/16/$31.00 ©2016 IEEE

degradation signals collected by multiple acceleration sensors under various operating conditions. Monitoring accuracy of the proposed method is also compared with that of a neural network-based method. Section V summarizes the research and draw conclusions.

II. MULTI-SENSOR DEGRADATION MONITORING BASED ON DBN

A. A Brief Introduction to DBN

Deep belief network is a kind of generative depth model which can describe the high order correlation of the input data, and has become a widely used deep learning structure both in research and application since it was firstly proposed by Hinton [7].

1) Deep belief network architecture

DBN consists of multiple hidden layers and one visible layer. As a kind of Bayes probability generation model, DBN is used to establish a joint distribution between the observation data with labels. The DBN is the stacked network of the restricted Boltzmann machine (RBM). Input parameters of DBN, which include total number of layers, preprocessed batch data, total number of hidden neurons in each hidden layer and maximum number of epochs, are initialized before the start of training.

For DBN classifier model with l hidden layers, the joint distribution between the visual variables and the l hidden layers $h^{(k)}(k=1,2,...,l)$ can be modeled, and the joint probability $p(v,h^{(1)},h^{(2)},...h^{(l)})$ of the whole DBN is:

$$p(v,h^{(1)},h^{(2)},...h^{(l)}) = p(v \mid h^{(1)})p(h^{(1)} \mid h^{(2)})\cdots p(h^{(l-1)},h^{(l)})$$

where $p(h^{(k)} \mid h^{(k+1)})$ is the factorial conditional distribution between the kth hidden layer and the $(k+1)th$ hidden layer.

Each layer of DBN is trained using the RBM learning rule. The positive learning phase transfers data from visible layer to hidden layer and determines the probability of generating hidden units as $p(h \mid v,W)$, whereas the negative learning phase extract a reconstruction of the previous visible layer and determines the probability of generating visible units as $p(v \mid h,W)$. Trained weights and generated visible units will be got after repetitive positive phase and negative phase in the training of RBM layers. The learning procedure function of generating a visible vector $p(v)$ by DBN learning process can be formulated using the probability $p(v \mid h,W)$ and the probability $p(h \mid v,W)$, as

$$p(v) = \sum_h p(h \mid v,W)\,p(v \mid h,W)$$

The DBN is trained layer by layer and the RBM is also individually trained, and the weights and biases are saved for further analyze. This individual training process is the pre-training of DBN classifier model, and is an unsupervised learning process. Reverse fine-tuning can be used to adjust all the parameters of DBN by using back-propagation training after pre-training.

2) DBN classifier model

Stacked RMB learning and back-propagation learning are the two steps of DBN training process. The following subsections describe these two steps in detail.

a) Stacked RBM

RBM is the development of Boltzmann machines (BM) which is proposed by D.H. Ackley et.al in 1985 [8], and were first introduced by Paul Smolensky in 1986 [9]. The network of RBM is fully connected between the two layers, yet no units within the same layer are connected to one another as shown in Fig. 1 In this way, the probability distribution function can be obtained easily [10]. The process of transforming data from visible layer to the hidden layer is finished through a sigmoid activation function based on the RBM learning rule [11].

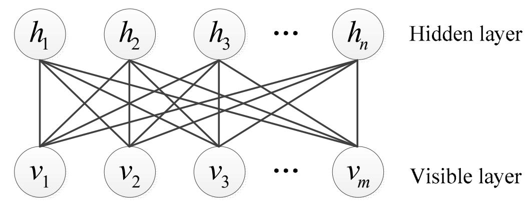

Figure 1. The structure of RBM

RBM is a special kind of Markov random field with 2 layer structure as shown in Fig. 1. RBM contains a visual layer v consists of m visible units $(v_1,v_2,...,v_m)$ and a hidden layer h consists of n hidden units $(h_1,h_2,...,h_n)$. Visible units usually obey the Bernoulli or Gaussian distribution and hidden units usually obey the Bernoulli distribution.

Energy function, which can be used to estimate the state of the RBM, always changes in the direction of decreasing, or staying in a fixed when the system evolves according to its internal power rules. It will tend to stable eventually. Once model parameters (i.e., weight w, visual layer bias b, hidden bias c) are given, the joint distribution $p(v,h;\theta)$ between visible vector v and hidden vector h can be formulated by energy function $E(v,h;\theta)$ expressed as

$$p(v,h;\theta) = \exp(-E(v,h;\theta))/Z$$

where $Z = \sum_v \sum_h \exp(-E(v,h;\theta))$ is the normalized factor.

The energy function of Bernoulli–Bernoulli RBM model is

$$E(v,h;\theta) = -\sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij}h_i v_j - \sum_{j=1}^{m} b_j v_j - \sum_{i=1}^{n} c_i h_i$$

where w_{ij} is the real weight between unit v_j and unit h_i, b_j and c_i are the real bias terms of the jth visual variable and the ith hidden variable respectively. Positive and negative phases of RBM learning can express mathematically as

$$p(h_i = 1 \mid v;\theta) = \sigma(\sum_{j=1}^{m} w_{ij}v_j + b_j)$$

$$p(v_j = 1 \mid h; \theta) = \sigma(\sum_{i=1}^{n} w_{ij} h_i + c_i)$$

where $\sigma(x) = \frac{1}{1+\exp(-x)}$, is the sigmoid function.

Similarly, energy function of Gaussian–Bernoulli RBM model is

$$E(v,h;\theta) = -\sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij} h_i v_j + \frac{1}{2}\sum_{j=1}^{m}(v_j - b_j)^2 - \sum_{i=1}^{n} c_i h_i$$

Positive and negative phases of RBM learning can express mathematically as

$$p(h_i = 1 \mid v; \theta) = \sigma(\sum_{j=1}^{m} w_{ij} v_j + b_j)$$

$$p(v_j = 1 \mid h; \theta) = N(\sum_{i=1}^{n} w_{ij} h_i + c_i, 1)$$

where v_j is the continuous value that obey Gaussian distribution.

Weight updating rules of RBM can be derived using the logarithm likelihood function gradient:

$$\Delta w_{ij} = E_{data}(v_j h_i) - E_{model}(v_j h_i)$$

where $E_{data}(v_j h_i)$ is the expectation for training set and $E_{model}(v_j h_i)$ is the expectation which defined in model.

Fig. 2 demonstrates the unsupervised DBN learning process.

The input data of the first RBM unit is visible layer. Then input data from RBM visible layer is transformed into hidden layer. The hidden layer of the first RBM is treated as the visible layer of the second RBM. Training process is continued for all the RBM until the training of DBN is accomplished.

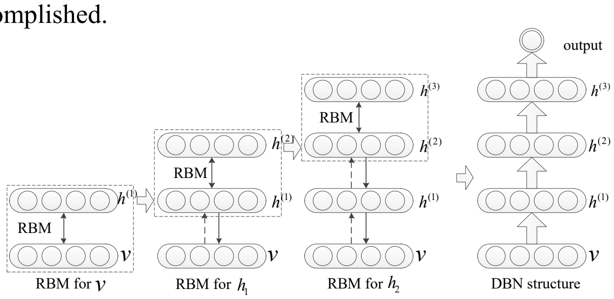

Figure 2. The training process of DBN

Trained RBMs are stacked to form DBN and trained weights and biases of RBM layers from unsupervised RBM learning are utilized in the DBN training process.

b) Back-propagation learning

The next step after layer by layer training process is the supervised learning process, which will be accomplished by the back-propagation (BP) training algorithm. This supervised learning process adjusts parameters in each layer to reduce the training error and improves the classification accuracy of the DBN classifier model. The back-propagation training process considers all DBN layers simultaneously. Labeled data are used in this process. Training error is calculated by comparing model outputs and target labeled data and all parameters in DBN classifier model are updated to minimize the training error [12]. The trained DBN classifier model can be further fine-tuned after supervised training process to improve the classification accuracy by fine-tuning algorithm.

Classification validity should be checked after the DBN training and fine tuning process. The ratio that number of misclassified points divided by total number of data points in the input dataset is served as the misclassification error. The validation process is carried out both in training and testing datasets. DBN model can be used in degradation monitoring based on the acquired sensor data when the misclassification error meets requirements.

B. Multi-sensor Data Fusion Technology

Multi-sensor data fusion is an information processing process by using the computer technology to analyze and synthesize the information of multiple sensors automatically in a certain principle, to complete the decision and estimation. Information fusion technology can be divided into three kinds of fusion methods namely data level fusion, feature level fusion and decision level fusion. Each fusion method has its advantage and disadvantage. The choice of fusion method is determined by the demand of user and types of sensors.

Single-sensor data is of huge uncertainty which will reduce the accuracy of classification and even appear error in monitoring complex equipment, hence multi-sensor data fusion technology show great superiority in condition monitoring for complex machine such as ball screw. Multi-sensor data fusion, which can provide degradation state information from different sides to improve the classifier accuracy and monitoring reliability, is an important development direction of artificial intelligence.

III. DEGRADATION CONDITION CLASSIFICATION METHOD FOR BALL SCREW

Multiple signals collected in different locations are needed to reflect the degradation condition of ball screw, which will increase the difficulty of feature extraction. Traditional methods of feature extraction and machine learning rely too much on signal processing technology and diagnosis experience, and usually have poor generalization ability. Aiming at these problems, an intelligent degradation condition classification method based on deep learning and multi-sensor data fusion is proposed in this paper. This method can adaptively extract degradation characteristics from raw signals and automatically distinguish the degradation degree. Fused spectra are used in this method as the input data of DBN classifier model.

The proposed method can be divided into four steps as shown in Fig. 3 data preprocessing, unsupervised pre-training, supervised training and degradation condition monitoring.

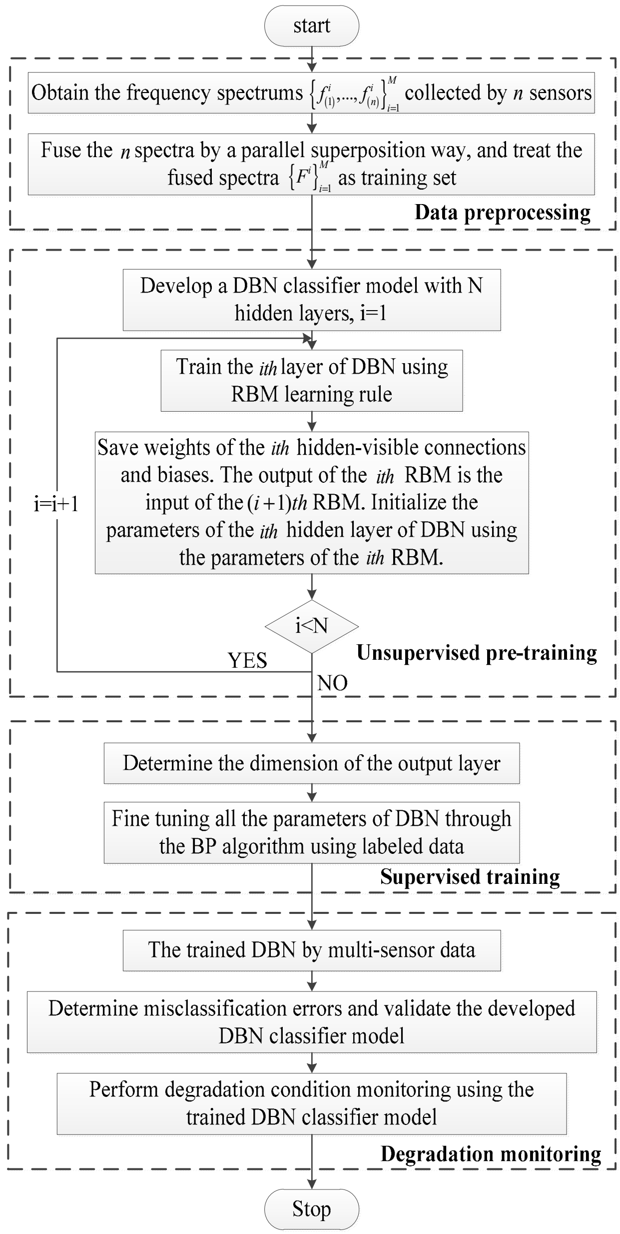

Figure 3. Flowchart of the proposed method

The first step is the sample acquisition and data preprocessing. After collecting the time domain signals from n different sensors, frequency spectra $\{f_{(1)}^i,...,f_{(n)}^i\}_{i=1}^M$ corresponding to these time domain signals are calculated, where M is the number of degradation sample in each group. Parallel superposition of the n spectra is done to increase the dimension of the feature space, and fused frequency spectra are directly as the input data of DBN and express as $\{F^i\}_{i=1}^M$, where $F^i = f_{(1)}^i \cup \cdots \cup f_{(n)}^i$.

The next step is the unsupervised pre-training. DBN classifier model with N hidden layers should be developed at first. The dimension of the input layer is equal to the sum of spectral dimensions of the n single sensor signals. The relation can express as $dim(F^i) = \sum_{j=1...n} dim(f_{(j)}^i)$, where dim represents spectral dimension. The unlabeled training sets $\{F^i\}_{i=1}^M$ are utilized to train DBN layer by layer using RBM learning rule. The output of the ith RBM is the input of the $(i+1)th$ RBM. Weights and biases of the ith RBM should be saved to initialize the parameters of the ith hidden layer of DBN. The pre-training of DBN is accomplished until the Nth RBM is trained.

The third step is the supervised training, which will be accomplished by the back-propagation training algorithm. The degradation grade corresponding to the dimension of the output layer should be determined. All the parameters of DBN are adjusted through the BP algorithm using labeled data. The target is to minimize the errors between the degradation condition grade labels and outputs calculated by the fused frequency spectra.

Finally, the misclassification error can be determined to validate the DBN classifier model after the DBN is trained by multi-sensor data. Then, the trained DBN classifier model can be used to monitor the degradation grade of the ball screw by collecting multi-sensor vibration signals.

Fused multi-sensor signals are more comprehensive, accurate and redundant compared with single-sensor signals. DBN mines essential characteristics from fused frequency spectra and establishes the non-liner mapping between fused frequency spectra and degradation grade. Hence, the proposed DBN-based degradation monitoring method could realize feature adaptive extraction and intelligent degradation monitoring.

IV. DEGRADATION MONITORING APPLICATIONS ON BALL SCREW

A. *Experiment Study Description*

Ball screw is a kind of precision transmission part in drive system of machine tool, and its performance degradation will lead to lower machining precision and even downtime which will result in huge economic losses. Therefore, it is important that monitoring the performance degradation degree of ball screw in improving machining stability and machining efficiency.

FFZD4010R-3 type of ball screw installed on the acceleration performance degradation test bench is chosen to research its degradation characteristics in this study. As shown in Fig. 4, this test bench is composed of ①driving motor, ② slide guide, ③ball screw, ④rack and pinion unit and ⑤ magnetic power brake. The test bench is designed to simulate the performance degradation process of ball screw during processing. Multiple acceleration sensors that reflect well on device status are used to collect vibration signals. The degradation degree of ball screw can be judged by analyzing these vibration signals. The object of this study is to classify different degradation conditions of ball screw based on the multi-sensor vibration signals.

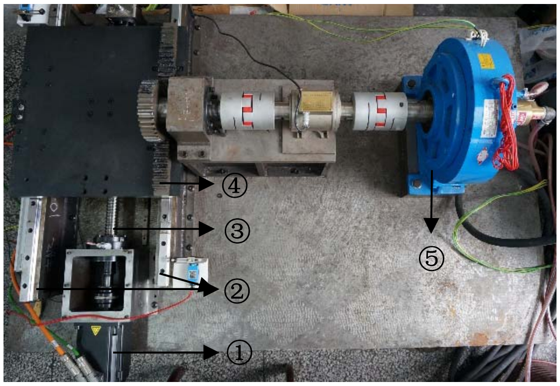

Figure 4. Degradation test bench for ball screw

B. Data Acquisition and Preprocessing

The test bench of ball screw simulates the whole performance degradation process of the ball screw from new to failure. The whole performance degradation process of ball screw is divided into 7 grades.

Vibration signals are collected in 7 degradation grades and 9 working conditions using 5 acceleration sensors. Nine working conditions are the cross combination of three speeds (100r/min, 300r/min, 800r/min) and three axial loads (0kN, 1kN, 2kN). Five groups of vibration signals are collected by 5 acceleration sensors installed on multiple positions and directions. Positions include ball nut and two bearing seats, and directions include axial and radial directions of screw. Five acceleration sensors collect signals at the same time. Each acceleration sensor collects 120 samples for the first 6 working conditions and 90 samples for the last 3 working conditions in each degradation condition. Hence 4950 samples are collected in each degradation condition, and 6930 samples are collected by each acceleration sensor for total 7 degradation grades. In conclusion, the degradation vibration signal of ball screw contains 34650 samples, which are collected in 9 working conditions and 7 degradation grades by 5 sensors installed on different positions and directions.

The next step is to preprocessing the collected degradation data. The spectrum of each sample, which is a vector of 512 lengths, is calculated respectively. So 5 groups of signals collected by 5 sensors produce 5 spectrum matrixes whose line is 6930 and column is 512. Then fuse the 5 spectrum matrixes according to parallel superposition way to obtain a spectrum matrix with 6930 lines and 2560 columns. Thus the number and the dimension of input data are 6930 and 2560 respectively, and this fused spectrum matrix is treated as the input of DBN. Degradation grade of ball screw is represented by 7-dimension vector whose nth value is equal to 1 whereas the other values are equal to 0 when the ball screw is in the nth degradation grade.

C. DBN Based Degradation Condition Monitoring

The trained DBN classifier model has one data layer on the bottom, one output layer on the top and two hidden layers whose neurons are 500 and 100 respectively. The neuron number of the input layer is equal to the dimension of the input spectrum dataset, and the neuron number of the output layer is also equal to the number of defined degradation grade of ball screw. The structure of DBN is 2560-500-100-7 in this study. The maximum numbers of training epochs used in this RBM training process and back propagation learning process are respectively set to 50 and 600, the learning rate is 0.1 and the batch size is 63. Weights of DBN are initialized randomly and biases are initialized to zero. In the trial, 70% of data samples are randomly selected for training, and the others are used for testing. Trial repeated twenty times.

Degradation condition classification results using the proposed method are shown in Fig. 5. The average training accuracy of these twenty trials is equal to 99.06% and the average testing accuracy is equal to 91.89%. Validation results indicate the effectiveness of the proposed method and mean that these seven degradation conditions of ball screw can be distinguished in a high accuracy. This high accuracy is mainly because that DBN has the capability of learning complex non-liner relationship between input data sets and different degradation states.

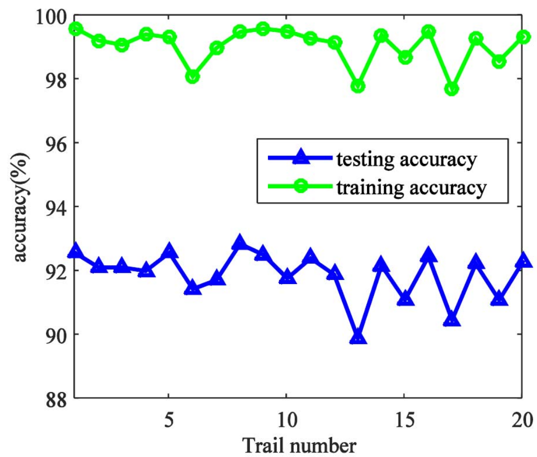

Figure 5. Classification results using the proposed method

The degradation grade of ball screw also classified using the neural network-based multi-sensor data fusion method for comparison. The trained BP neural network has one input layer, one output layer and two hidden layers. Trials also repeated twenty times. Degradation classification results using the BP neural network are shown in Fig. 6. The average training accuracy of these twenty trials is equal to 91.91% and the average testing accuracy is equal to 80.82%. It can be seen that the generalization ability and classification accuracies of neural network-based multi-sensor data fusion method are obviously poorer than the proposed method. More seriously, the fifth trial result fall into local optimum as shown in Fig. 6, and training accuracy and testing accuracy reach to 67.23% and 51.32% respectively, which are much lower than requirements. Local optimum greatly reduces the stability of classification. Otherwise, the standard deviation of the proposed method is equal to 0.75%, whereas the standard deviation of the BPNN method is equal to 7.55%. Average accuracies are larger and the standard deviation is much smaller compared with the neural network-based multi-sensor data fusion method, which means the proposed method is more effective and stable than the neural network-based multi-sensor data fusion method and has more superiority.

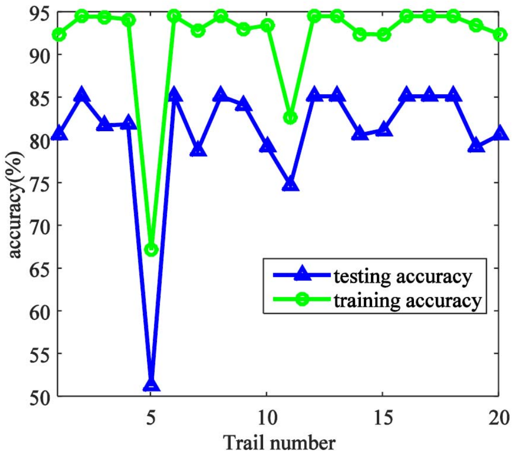

Figure 6. Classification results using the BPNN based method

V. CONCLUSIONS

This paper presented a novel degradation monitoring method for ball screw by utilizing DBN and multi-sensor data fusion technology. The DBN based degradation monitoring method can be structured in four consecutive stages: (1) data acquisition and calculating the fused frequency spectra by parallel superposition; (2) pre-training the DBN classifier model using stacked RBM; (3) adjusting parameters of pre-trained DBN using labeled data by supervised BP algorithm; (4) monitoring the degradation grade of ball screw. The effectiveness of the proposed method was validated by calculating classification accuracies using the fused spectra datasets that contain massive samples of ball screw. Monitoring results of the DBN-based classifier model were compared with that of the neural network-based method, and the proposed method was proved to have higher accuracy and stability. These study results indicated that DBN-based classifier model has better degradation condition monitoring performance and more superiority.

REFERENCES

[1] R. J. Kuo, P. H. Cohen, "Multi-sensor integration for on-line tool wear estimation through radial basis function networks and fuzzy neural network," Neural Networks, vol.12, No.2, pp.355-370, Mar, 1999.

[2] M. S. Safizadeh, S. K. Latifi, "Using multi-sensor data fusion for vibration fault diagnosis of rolling element bearings by accelerometer and load cell," Information Fusion, vol.18, pp.1-8, Jul, 2014.

[3] G. E. Hinton, R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," Science, vol.313, No.5786, pp.504-507, Jul, 2006.

[4] P. Tamilselvan, P. Wang, "Failure diagnosis using deep belief learning based health state classification," Reliability Engineering & System Safety, vol.115, pp.124-135, Jul, 2013.

[5] V. T. Tran, F. AlThobiani, and A. Ball, "An approach to fault diagnosis of reciprocating compressor valves using Teager–Kaiser energy operator and deep belief networks," Expert Systems with Applications, vol.41, No.9, pp.4113-4122, Jul, 2014.

[6] F Jia, Y Lei, J Lin, and N Lu, "Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data," Machine System and Signal Processing, vol.72-73, pp.303-315, May, 2016.

[7] G. E. Hinton, S. Osindero, and Y. W. Teh, "A fast learning algorithm for deep belief nets," Neural computation, vol.18, No.7, pp.1527-1554, Jul, 2006.

[8] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, "A learning algorithmfor Boltzmann machines," Cognitive Science, vol.9, No.1, pp.147-169, Jan-Mar, 1985.

[9] P. Smolensky, "Chapter 6: Information Processing in Dynamical Systems:Foundations of Harmony Theory. In D. E. Rumelhart, Parallel Distributed Processing: Explorations in the Microstructure of Cognition," Cambridge, MA: MIT Press, 1986, pp.194 – 281.

[10] D. Yu, L. Deng, "Deep learning and its applications tosignal and information processing," IEEE SignalProcessing Magazine, vol.28, No.1, pp.145-149, Jan, 2011.

[11] G. Hinton, "A practical guide to training restricted Boltzmann machines," Momentum, vol.9, No.1, pp.926, 2010.

[12] Y. Bengio, P. Lamblin, D. Popovici, and H Larochelle, "Greedy layer-wise training of deep networks," Advances in Neural Information Processing Systems, 2007, pp.153.